DOCUMENT RESUME

ED 426 656 HE 031 805

AUTHOR Copeland, H. Liesel; Hewson, Mariana
TITLE Clinical Teaching Effectiveness Instrument: Development and Psychometric Testing.
PUB DATE 1999-04-00
NOTE 8p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, April 19-23, 1999). For a related paper, see HE 031 804.
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Clinical Teaching (Health Professions); Evaluation Methods; *Faculty Evaluation; Feasibility Studies; *Graduate Medical Education; Higher Education; *Medical Education; Program Evaluation; Psychometrics; Rating Scales; *Teacher Effectiveness; Test Reliability; Test Validity
IDENTIFIERS *Cleveland Clinic OH

ABSTRACT

This report describes the development and psychometric qualities of a new instrument to assess clinical teaching effectiveness in medical education. The strength of the instrument is seen to lie in the qualitative development process involving iterative checking with key stakeholders; its high reliability, validity, and feasibility; and its ease of implementation within a coherent institution-wide feedback system. The instrument was developed in conjunction with current literature and with data collected from a series of interviews with relevant stakeholders. The instrument has 15 rating items, and one general item asking residents if they would recommend this individual as a clinical teacher. The instrument was implemented in 1997-98 across all 41 clinical departments of the Cleveland (Ohio) Clinic Academic Medical Center. This report presents psychometric data on instrument characteristics, modifying variables, reliability, content validity, criterion-related validity, and feasibility/usefulness. The instrument is found to be potentially useful in instructor evaluation, research on variables affecting clinical teaching, and staff development. (DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 426 656

Clinical Teaching Effectiveness Instrument: Development and Psychometric Testing

Presented at the American Educational Research Association Annual Conference,
April 1999, Montreal

H Liesel Copeland PhD & Mariana Hewson PhD
Education Division TT31
Cleveland Clinic Foundation
9500 Euclid Ave., Cleveland, OH 44195

BEST COPY AVAILABLE

HE 031 205

U.S. DEPARTMENT OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Liesel Copeland

Clinical Teaching Effectiveness Instrument:
Development and Psychometric Testing

H Liesel Copeland, Ph.D. & Mariana Hewson, Ph.D.

We developed a new clinical teaching effectiveness instrument that was theory-based and generic across the entire medical center for the purpose of improving teaching competencies. Our aim is to provide clinician-educators with regular feedback on their teaching performance in order to enable them to improve their teaching effectiveness. Our purpose is to report on the development and psychometric qualities of this new instrument. The strength of our instrument lies in the qualitative development process involving iterative checking with key stakeholders, its high reliability, validity and feasibility, and its implementation within a coherent institution-wide feedback system.

Purpose

Our Academic Medical Center is committed to high quality medical education and is responsible to our stakeholders to demonstrate the effectiveness of our clinical teachers. It is also committed to improving the teaching abilities of all our clinician-educators. Ratings from students are commonly used and are considered to be an essential component of teaching evaluation systems in tertiary and professional educational institutions. A review of the literature (1) confirmed that: student ratings are reliable; can be correlated with measures such as student learning, instructor self-evaluations, and peer ratings; and are generalizable across different teaching situations. By providing clinician-educators with ongoing feedback on their teaching performance, especially in the context of specific teaching standards, we can enable them to adjust their teaching behaviors and improve their teaching effectiveness. The feedback can also be useful in making decisions about academic promotion as well as the allocation of teaching responsibilities within departments.

Our previous evaluation of clinical teaching effectiveness involved diverse, department-specific instruments that lacked comparability. We needed a new instrument that was theory-based and generic across the institution for the purpose of comparing teaching competencies amongst faculty, departments, and divisions[1]. The key qualities required for the new instrument were that it should be practical and feasible (eg, short, visually appealing, and scannable), useful for clinician-educators in motivating self-improvement and for the annual performance review, clinically credible for all divisions, valid, and reliable. The purpose of this paper is to report on the development and psychometric qualities of this new clinical teaching evaluation instrument.

Methods

The Clinical Teaching Effectiveness instrument was developed in conjunction with current literature (2-8) and through data collected from a series of interviews with all relevant stakeholders using qualitative methods. The first prototype of the instrument was based on an inventory of effective clinical teaching behaviors (3), which was consistent with Hewson's model of tailored clinical teaching (4). The instrument was first drafted by a committee in the department of medicine (composed of the residency program director, education administrator, chief resident, a community physician, and medical educator). The draft instrument was then modified in an iterative process through numerous meetings with stakeholder representatives from each of the groups (residency and medical student program directors, department and division chairs, educational administrators, clinician-educators, and residents) and from the major clinical teaching divisions (medicine, pediatrics, psychiatry, surgery, anesthesiology, radiology and pathology). When the process of continual modification and refinement reached the "point of redundancy" (ie, the meetings no longer resulted in new ideas or disagreements), we concluded that a type of serial consensus had been attained, and we finalized the instrument. This iterative process allowed us to obtain "buy in" from all areas within the institution, and helped us inform people of the impending changes in the evaluation system.

The new Clinical Teaching Effectiveness instrument has fifteen rating items plus one general item asking residents if they would recommend this staff as a clinical teacher (Table 1). There is space for comments. Each rating item uses a five point evaluation scale where 1 = Never/ Poor Teacher and 5 = Always/ Superb Teacher. Resident and student evaluators are guaranteed anonymity.

In order to check modifying variables, we collect demographic information on the time spent with particular clinician-educators, their residency program and their level of training. This provides an opportunity to research their effect.

---

[1]At our institution, departments are subordinate to divisions.

In 1997-8 we implemented the new instrument across all departments and we report here on data collected for each of 41 clinical departments. Data from the instrument are systematically fed back to individuals, program directors, and department and division chairs. All data for each clinician-educator are explicitly reviewed in the annual performance review. Reliability estimates were computed using Genova and other statistical computations were obtained through SPSS for Windows.

Table 1

Offers regular feedback (positive & negative) ($\bar{x}$=4.03, .94)	Teaches effective patient/family communication skills ($\bar{x}$=4.08, .95)	Clearly specifies what I am expected to know and do ($\bar{x}$=3.94, .96)
Establishes a good learning environment ($\bar{x}$=4.28, sd=87)	Teaches principles of cost-appropriate care ($\bar{x}$=3.94, sd=.96)	Gives clear explanations for opinions, advice, actions ($\bar{x}$=4.23, .89)
Observes and coaches my clinical/ technical skills ($\bar{x}$=3.92 (1.0)	Stimulates me to learn independently ($\bar{x}$=4.18, sd=.85)	Allows me appropriate autonomy ($\bar{x}$=4.18, sd=.92)
Teaches diagnostic skills ($\bar{x}$=4.19, sd=.85)	Adjusts teaching to my needs ($\bar{x}$=4.13, sd=.92)	Organizes time for teaching and care-giving ($\bar{x}$=4.08, sd=.97)
Incorporates research data & practice guidelines ($\bar{x}$=4.17, sd=.90)	Asks questions that promote learning ($\bar{x}$=4.20, sd=.91)	Provides effective teaching at multiple sites ($\bar{x}$=4.21, sd=.90)

Results

Instrument Characteristics Instruments were completed by medical students, residents, and fellows. An average of seven instruments per faculty were collected for 570 faculty (a total of 3827 instruments, plus 276 left blank due to self-reported insufficient time with faculty member). The average rating for all fifteen items is 4.11 (sd=.76) with mean ratings for individual items ranging from 3.92 to 4.28 (see Table 1). Skipped items usually occurred under 5% of the time, with 4 under 12% and one (communication) under 17%. Though the distribution of the average scores is negatively skewed, it approximates a normal distribution. All the items on the instrument are inter-related with correlations ranging from .57 to .76. A factor analysis of the fifteen rating items resulted in a single component explaining 69% of the variance, indicating we are measuring one core concept. All fifteen rating items loaded with at least .79. The highest loading items were: 1) adjusts teaching to my needs, 2) provides effective teaching at multiple sites, 3) stimulates me to learn independently, 4) teaches diagnostic skills, and 5) asks questions that promote learning.

Modifying Variables The average time a trainee spent with a faculty member was six weeks. Using ANOVA (through GLM in SPSS), we found no statistically significant effect on ratings for the time spent with the clinician-educator nor for the trainee level. Though no overall difference was found for trainees, a trend is apparent. Analyzing post-hoc tests we find medical students ($\bar{x}$=4.38, 95% ci: 4.07-4.68) rate faculty significantly higher than residents ($\bar{x}$=4.07, 95%ci: 4.03-4.11) or fellows ($\bar{x}$=4.03, 95% ci: 3.96-4.10).

Reliability: The reliability of this instrument was estimated through generalizability analysis with computation of a g-coefficient. We entered three sources of differences in scores (effects) into the analysis: 1) the clinician-educators, 2) the items, and 3) the trainees (raters) who were nested within clinician-educators (ie, every clinician-educator was rated by a different trainee). When computing the g-coefficient, items were fixed at 15 and raters were considered random. Variance estimates were obtained from a data set of 295 clinician-educators, each evaluated by 5 trainees. Variance estimates for the components of this study ranged from .048 (raters nested in faculty) to .773 (item-rater

interaction). The g-coefficient (reliability coefficient) for our design is .935; even if we were to use one rater the g-coefficient would be .742 and with seven raters it rises to .953, indicating our instrument is highly reliable. This also means that the 95% confidence interval for the mean is ± .377 for 5 raters (± .752 if 1 rater was used). The high inter-correlations and the factor analysis show that this instrument has a high internal consistency. This is confirmed by our computation of coefficient alpha being .958.

Validity:

Content: A content validation study analyzes whether the items on the instrument adequately represent the domain of interest. We performed a modified content (face) validation study by comparing our instrument with several alternative clinical teaching evaluation instruments. Of the fifteen items on the University of Toronto's faculty teaching instrument (9), eleven were represented on our newly developed instrument (73%). Of the eighteen items on Westberg & Jason's sample instrument, nine were embodied in our instrument's items (50%). Of the twenty-three items on our most used former teaching instrument, eleven are represented on the new instrument (48%). We also assessed validity and comprehensiveness by analyzing, during item development, concept congruence in our data sources. Complete congruence was obtained for: 1) offers feedback, 2) establishes a good learning climate, 3) observes and coaches clinical/technical skills, 4) teaches medical knowledge (diagnostic skills, research data and practice guidelines, communication skills, cost-appropriate care), and 5) stimulates independent learning. Concepts that were common in the literature but infrequently mentioned by residents were: 1) adjusts teaching to learner's needs, 2) asks questions to actively involve learners, 3) specifies expectations, and 4) gives clear explanations and answers questions. Concepts mentioned by most stakeholders but less commonly in the literature are: 1) provides autonomy, 2) organizes time for teaching and care-giving, and 3) provides training in multiple sites.

Criterion-related Criterion-related validation studies assess the relationship between scores on the instrument and some criterion measure. We used averaged 1996-7 scores on a former instrument from five divisions as a "retrospective" criterion. Table 2 shows the correlations between clinician-educators' scores on the current and old instrument showing that the validity is good (ie, the instrument is practical and useful) (1). This indicates that a fundamental criterion of teaching is being assessed and that the new instrument is providing more specific information.

Table 2

	n	mean (sd)	correlation with new instrument	p (n)
Average of all old instrument items	421	4.08 (.5587)	.428	<.01 (351)
Old instrument "overall" item average	420	4.09 (.6496)	.433	<.01 (350)
Average of new instrument items	570	4.11 (.5167)	-	-

A second criterion involved our institutional Alumni Survey data where alumni (3 years post residency) named specific clinician-educators for excellent teaching (they were recalling educators from 1992-5). We selected the top 16% of current clinician-educators (those with scores 1 standard deviation above the mean on the Clinical Teaching Effectiveness instrument) and compared these individuals with those named as excellent teachers by alumni. Of the top-rated clinician-educators on the new instrument, we only compared those who had been appointed prior to 1996. Looking at these top-rated clinician-educators, 41.4% (24 of 58) are also mentioned by alumni.

Feasibility and Usefulness The new instrument is highly useable, though we are not using the scanner due to incompatibility in saving verbatim comments. Based on reports from individual clinical-educators, the new Clinical Teaching Effectiveness instrument is raising awareness of specific clinical teaching behaviors and more people are seeking help with their teaching.

Discussion

Though a trainee's ratings of faculty is a highly valued component of teaching evaluation, it is advisable to gather multiple sources of data for a complete evaluation of teacher effectiveness. Alternative sources include peer evaluations, self evaluations, or observations. A secondary source of data for decisions on teaching effectiveness would be beneficial. The high mean and slight skewness of our data does imply that we have a ceiling effect and despite being able to differentiate among high and low teachers we are unable to discriminate amongst our highly competent teachers. This is not a large concern since our aim is to ensure all faculty reach a specified level of effectiveness and to help those who have not achieved this level. We have studied the response rate by item and find none is so low as to jeopardize interpretation.

Conclusion

The Clinical Teaching Effectiveness instrument has now been used in all the divisions at our Academic Medical Center and was determined to be reliable and valid, as well as feasible and potentially (at this point) highly useful. The items represent specific theoretical constructs important to clinical teaching and are therefore useful in promoting self-improvements among our faculty. The strength of our instrument lies in the qualitative development process where iterative checking with key stakeholders and informants occurred. Furthermore, we are now able to provide a thorough explanation of and justification for our measure of teaching effectiveness. We can now address research questions concerning variables affecting clinical teaching and can compare the teaching of individuals and different departments. We are also able to give guidance for interpretation of a score by providing confidence intervals. By providing a well documented and theoretically based instrument we can not only improve the teaching at this medical center but also promote the importance of clinical teaching and demonstrate the institutional value placed on such efforts.

References:

(1) Pritchard, RD, Watson, MD, Kelly, K, and Paquin, AR (1998). Helping Teachers Teach Well. San Francisco: The New Lexington Press.

(2) Anderson, D. C., Harris, I. B., Allen, S., Satran, L., Bland, C. J., Davis-Feickert, J. A., Poland, G. A., and Miller, W. J. (1991). Comparing student's feedback about clinical instruction with their performances. *Academic Medicine*, 66(1), 29-34.

(3) Hewson, M G. and Jensen, N. M. (1990). An inventory to improve clinical teaching in the general internal medicine clinic. *Medical Education*, 24, p. 518-27.

(4) Hewson. M. (1992). Clinical Teaching in the Ambulatory Setting. *Journal of General Internal Medicine*, 7, p. 76-82.

(5) Irby, D. M. (1994). What clinical teachers in medicine need to know. *Academic Medicine*. 69(5), p.333-42.

(6) Skeff, K. M., Stratos, G. A., and Bergen, M. R., (1992). Evaluation of a medical faculty development program: a comparison of traditional pre/post and retrospective pre/post self assessment ratings. *Evaluation & The Health Professions*, 15(3), 350-66.

(7) Westberg, J., Jason, H. (1993). Collaborative Clinical Education: The Foundation of Effective Health Care, New York: Springer Publishing.

(8) Whitman, N. and Lawrence, P. (1991). *Surgical Teaching: Practice makes perfect*. Department of Family and Preventive Medicine, University of Utah School of Medicine.

(9) Slade, S. (1993). *Evaluation of Faculty Teaching*. Department of Family & Community Medicine, University of Toronto.